

CLAIMS

1. A process for real time analysis of text and/or media content and relating
5 information to the content, comprising the steps of:

analyzing said content in real time;

wherein said analyzing step analyzes said content for semantic and
conceptual use;

providing a set of informational documents;

10 wherein said informational documents comprise any of text, Web, and media
documents;

providing a pre-processed analysis of said informational documents;

wherein said pre-processed analysis is an analysis of said informational
documents for semantic and conceptual use;

15 identifying informational documents related to said analyzed content using
said pre-processed analysis;

providing a user with a description of each identified informational
document;

accepting user input for selecting an identified informational document; and

20 displaying the selected identified informational document to the user.

2. The process of Claim 1, wherein said identifying step identifies related
informational documents by finding informational documents that are similar in
words, semantically or conceptually, to the analyzed content.

25 3. The process of Claim 1, further comprising the step of:

storing descriptors for each informational document.

retrieving descriptions of each identified informational document from said
stored descriptors.

30 4. The process of Claim 1, wherein said set of informational documents are
stored in a central storage device.

5. The process of Claim 1, wherein said pre-processed analysis creates a list of words and calculates the frequency that the words appear in said set of informational documents.

5 6. The process of Claim 5, wherein said pre-processed analysis translates similar words into the same word.

7. The process of Claim 1, wherein said pre-processed analysis generates collocations of words that appear together and calculates the frequency of pairs of words and the frequency of the words appearing together in said informational documents.

10

8. The process of Claim 7, wherein said pre-processed analysis finds relations between collocations to learn their meaning/context.

15

9. The process of Claim 1, wherein said pre-processed analysis uses a signature algorithm to calculate signatures for blocks of text, wherein a signature is a vector of words and their weighting within an informational document; wherein the weighting is determined by the importance of a word in the collocations and within the document.

20

10. The process of Claim 9, wherein said pre-processed analysis calculates signatures for Web pages, text tags associated with images, and blocks of text.

25 11. The process of Claim 9, wherein said pre-processed analysis creates an index for each word from a signature vector for an informational document and saves the index, word, text document, and weight of the word into a database that is used to find text documents that have similar signatures.

30 12. The process of Claim 9, wherein said pre-processed analysis uses the signatures and weights of the words to create sets of documents that have similar signatures.

13. The process of Claim 1, further comprising the step of:
collecting text documents and multimedia from Web pages across the Internet using a Web crawler and placing them into said set of informational documents.

5

14. A process for real time analysis of text and/or media content in a workflow application and relating information to the content, comprising the steps of:

automatically analyzing said content in real time as said content is being entered or reviewed by a user;

10 wherein said analyzing step analyzes said content for semantic and conceptual use;

providing a set of informational documents;

wherein said informational documents comprise any of text, Web, and media documents;

15 providing a pre-processed analysis of said informational documents;

wherein said pre-processed analysis is an analysis of said informational documents for semantic and conceptual use;

identifying informational documents related to said analyzed content using said pre-processed analysis;

20 wherein said identifying step identifies related informational documents by finding informational documents that are similar in words, semantically or conceptually, to the analyzed content;

providing a user with a description of each identified informational document;

25 accepting user input for selecting an identified informational document; and displaying the selected identified informational document to the user.

15. A process for real time analysis of media content and relating information to the content, comprising the steps of:

30 extracting metadata from said media content in real time as said content is being viewed by a user;

providing a set of informational documents;

wherein said informational documents comprise any of text, Web, and media documents;

providing a pre-processed analysis of said informational documents;

5 wherein said pre-processed analysis is an analysis of said informational documents for semantic and conceptual use;

identifying informational documents related to said metadata using said pre-processed analysis;

10 wherein said identifying step identifies related informational documents by finding informational documents that are similar in words, semantically or conceptually, to said metadata;

providing a user with a description of each identified informational document;

accepting user input for selecting an identified informational document; and displaying the selected identified informational document to the user.

15

16. The process of Claim 15, wherein a broadcaster provides customized informational documents and specifies their relevance to be used by said identifying step.

20 17. The process of Claim 15, wherein a producer of said media content provides customized informational documents and specifies their relevance to be used by said identifying step.

25 18. The process of Claim 15, wherein said extracting step creates metadata for said media content by analyzing said media content if said media content does not have associated in-band metadata.

19. An apparatus for real time analysis of text and/or media content and relating information to the content, comprising:

30 a module for analyzing said content in real time;

wherein said analyzing module analyzes said content for semantic and conceptual use;

a set of informational documents;

wherein said informational documents comprise any of text, Web, and media documents;

a pre-processed analysis of said informational documents;

wherein said pre-processed analysis is an analysis of said informational documents for semantic and conceptual use;

a module for identifying informational documents related to said analyzed content using said pre-processed analysis;

a module for providing a user with a description of each identified informational document;

a module for accepting user input for selecting an identified informational document; and

a module for displaying the selected identified informational document to the user.

20. The apparatus of Claim 19, wherein said identifying module identifies related informational documents by finding informational documents that are similar in words, semantically or conceptually, to the analyzed content.

21. The apparatus of Claim 19, further comprising:

a module for storing descriptors for each informational document.

a module for retrieving descriptions of each identified informational document from said stored descriptors.

22. The apparatus of Claim 19, wherein said set of informational documents are stored in a central storage device.

23. The apparatus of Claim 19, wherein said pre-processed analysis creates a list of words and calculates the frequency that the words appear in said set of informational documents.

24. The apparatus of Claim 23, wherein said pre-processed analysis translates similar words into the same word.

25. The apparatus of Claim 19, wherein said pre-processed analysis generates collocations of words that appear together and calculates the frequency of pairs of words and the frequency of the words appearing together in said informational documents.

5

26. The apparatus of Claim 25, wherein said pre-processed analysis finds relations between collocations to learn their meaning/context.

10 27. The apparatus of Claim 19, wherein said pre-processed analysis uses a signature algorithm to calculate signatures for blocks of text, wherein a signature is a vector of words and their weighting within an informational document; wherein the weighting is determined by the importance of a word in the collocations and within the document.

15 28. The apparatus of Claim 27, wherein said pre-processed analysis calculates signatures for Web pages, text tags associated with images, and blocks of text.

20 29. The apparatus of Claim 27, wherein said pre-processed analysis creates an index for each word from a signature vector for an informational document and saves the index, word, text document, and weight of the word into a database that is used to find text documents that have similar signatures.

25 30. The apparatus of Claim 27, wherein said pre-processed analysis uses the signatures and weights of the words to create sets of documents that have similar signatures.

31. The apparatus of Claim 19, further comprising:
a module for collecting text documents and multimedia from Web pages across the Internet using a Web crawler and placing them into said set of
30 informational documents.

32. An apparatus for real time analysis of text and/or media content in a workflow application and relating information to the content, comprising:

a module for automatically analyzing said content in real time as said content is being entered or reviewed by a user;

wherein said analyzing module analyzes said content for semantic and conceptual use;

5 a set of informational documents;

wherein said informational documents comprise any of text, Web, and media documents;

a pre-processed analysis of said informational documents;

10 wherein said pre-processed analysis is an analysis of said informational documents for semantic and conceptual use;

a module for identifying informational documents related to said analyzed content using said pre-processed analysis;

15 wherein said identifying module identifies related informational documents by finding informational documents that are similar in words, semantically or conceptually, to the analyzed content;

a module for providing a user with a description of each identified informational document;

a module for accepting user input for selecting an identified informational document; and

20 a module for displaying the selected identified informational document to the user.

33. An apparatus for real time analysis of media content and relating information to the content, comprising:

25 a module for extracting metadata from said media content in real time as said content is being viewed by a user;

a set of informational documents;

wherein said informational documents comprise any of text, Web, and media documents;

30 a pre-processed analysis of said informational documents;

wherein said pre-processed analysis is an analysis of said informational documents for semantic and conceptual use;

a module for identifying informational documents related to said metadata using said pre-processed analysis;

wherein said identifying module identifies related informational documents by finding informational documents that are similar in words, semantically or conceptually, to said metadata;

a module for providing a user with a description of each identified informational document;

a module for accepting user input for selecting an identified informational document; and

a module for displaying the selected identified informational document to the user.

34. The apparatus of Claim 33, wherein a broadcaster provides customized informational documents and specifies their relevance to be used by said identifying step.

35. The apparatus of Claim 33, wherein a producer of said media content provides customized informational documents and specifies their relevance to be used by said identifying step.

36. The apparatus of Claim 33, wherein said extracting step creates metadata for said media content by analyzing said media content if said media content does not have associated in-band metadata.